arXiv:quant-ph/9604001v1 1 Apr 1996

Mathematical Techniques for Quantum Communication Theory*

Christopher A. Fuchs and Carlton M. Caves

Center for Advanced Studies, Department of Physics and Astronomy,

University of New Mexico, Albuquerque, NM 87131–1156

Abstract

We present mathematical techniques for addressing two closely related questions in quantum communication theory. In particular, we give a statistically motivated derivation of the Bures-Uhlmann measure of distinguishability for density operators, and we present a simplified proof of the Holevo upper bound to the mutual information of quantum communication channels. Both derivations give rise to novel quantum measurements.

1 Introduction

Suppose a quantum system is secretly prepared in one of two known, but non-orthogonal—or even mixed—quantum states $\hat{\rho}_0$ and $\hat{\rho}_1$. Because of the fundamental indeterminism of quantum mechanics, there is no way to discern reliably via measurement which of the two states has actually been prepared. One can still ask, however, which measurement among all possible quantum measurements will have an outcome that most likely distinguishes the one preparation from the other? Or, which measurement will gather the most Shannon information about the preparation if prior probabilities for the preparations are at hand? These questions, though not identical, are typical of quantum communication theory and contain to some extent the same mathematical difficulties. Here we develop mathematical techniques for addressing both questions.

In Section II we tackle a particular version of the first question by giving a statistically motivated derivation of the Bures-Uhlmann [1, 2, 3] measure of distinguishability for density operators and exploring the new quantum measurement that thus appears. In Section III we make progress toward the second question by simplifying the derivation of the Holevo upper bound [4, 5, 6, 7] on the maximum mutual information for binary quantum communication channels; by way of this, we find a measurement that often comes close to attaining the actual maximum value.

2 Statistical Distinguishability and Fidelity

Consider two distinct probability distributions p_{0b} and p_{1b} ($b = 1, \ldots, N$) for an experiment with N outcomes. Two common measures of the statistical distinguishability of these distributions are the Kullback-Leibler divergence or relative information [8, 9, 10],

$$K(p_0/p_1) = \sum_{b=1}^{N} p_{0b} \ln\left(\frac{p_{0b}}{p_{1b}}\right) \ , \tag{1}$$

and the Bhattacharyya-Wootters distance [9, 11, 12],

$$B(p_0, p_1) = \cos^{-1}\left(\sum_{b=1}^{N} \sqrt{p_{0b}}\sqrt{p_{1b}}\right) \ . \tag{2}$$

Both of these quantities take on a minimum value of zero if and only if the distributions are not distinguishable at all, i.e., $p_{0b} = p_{1b}$ for all b, but they define different notions of maximal distinguishability. A point of similarity between these measures is that when $p_{1b} = p_{0b} + \delta p_b$, to lowest order both are proportional to powers of the Fisher information [12, 13, 14, 15] at the point p_0 in the probability simplex,

$$ds^2 = \sum_{b=1}^{N} \frac{(\delta p_b)^2}{p_{0b}} \ . \tag{3}$$

This quantity places the ultimate limit on convergence in maximum likelihood parameter estimation [10, 16] and has recently found quite a use itself within the quantum context [17, 18, 19]. If the probability simplex is thought of as a Riemannian manifold with line element given by Eq. (3), the Bhattacharyya-Wootters distance is just the geodesic distance between the points p_0 and p_1 [12, 14].

The problem of statistically distinguishing the states $\hat{\rho}_0$ and $\hat{\rho}_1$ via a quantum measurement boils down to using a measurement with N outcomes (though N can be arbitrary) to generate the probability distributions p_0 and p_1 used in the measures (1) and (2). The optimal quantum measurement with respect to either the Kullback-Leibler or Bhattacharyya-Wootters distinguishability measure is just that measurement which makes either of the respective quantities as large as it can possibly be.

These ideas are made precise through a formalization of the most general measurements allowed by quantum theory, the positive-operator-valued measures (POVM) [20]. A POVM is a set of non-negative, Hermitian operators $\hat{E}_b$ which are complete in the sense that $\sum_b \hat{E}_b = \hat{\mathbb{1}} =$ (identity operator). The subscript b here, as before, indexes the possible outcomes of the measurement. Naturally, the conditions on the $\hat{E}_b$ are those necessary and sufficient for the standard expression $\mathrm{tr}(\hat{\rho}\hat{E}_b)$ to be a valid probability distribution for the b. The optimal distinguishability measurements $\{\hat{E}_b^{\mathrm{K}}\}$ and $\{\hat{E}_b^{\mathrm{B}}\}$ for the states $\hat{\rho}_0$ and $\hat{\rho}_1$ with respect to the measures (1) and (2) are just those which attain

$$K(\hat{\rho}_0/\hat{\rho}_1) \equiv \max_{\{\hat{E}_b\}} \sum_b \Big(\mathrm{tr}\hat{\rho}_0\hat{E}_b\Big) \ln\left(\frac{\mathrm{tr}\hat{\rho}_0\hat{E}_b}{\mathrm{tr}\hat{\rho}_1\hat{E}_b}\right) \ , \tag{4}$$

and

$$B(\hat{\rho}_0, \hat{\rho}_1) \equiv \max_{\{\hat{E}_b\}} \cos^{-1}\left(\sum_b \sqrt{\mathrm{tr}\hat{\rho}_0\hat{E}_b}\sqrt{\mathrm{tr}\hat{\rho}_1\hat{E}_b} \right) \ . \tag{5}$$

Notice again that the number N of measurement outcomes in these definitions has not been fixed at the outset as it is in the classical expressions (1) and (2).

The difficulty that crops up in extremizing quantities like Eqs. (4) and (5) is that, so far at least, there seems to be no way to make the problem amenable to a variational approach: the problems associated with allowing N to be arbitrary while enforcing the constraints on positivity and completeness for the $\hat{E}_b$ appear to be intractable. New methods are required. Fortunately, the Bhattacharyya-Wootters distinguishability measure appears to be "algebraic" enough that one might well imagine using standard operator inequalities, such as the Schwarz inequality for operator inner products, to aid in finding an explicit expression for $B(\hat{\rho}_0, \hat{\rho}_1)$. Progress toward finding a useful expression for $K(\hat{\rho}_0/\hat{\rho}_1)$ will, for just this reason, be impeded by the "transcendental" character of the logarithm in Eq. (4).

At this juncture we turn our focus to optimizing the Bhattacharyya-Wootters measure of distinguishability over all quantum measurements. For simplicity, here and throughout the remainder of the paper, we assume the density operators $\hat{\rho}_0$ and $\hat{\rho}_1$ to be finite dimensional and invertible. In this case, we shall show that

$$B(\hat{\rho}_0, \hat{\rho}_1) = \cos^{-1}\left(\mathrm{tr}\sqrt{\hat{\rho}_1^{1/2}\hat{\rho}_0\hat{\rho}_1^{1/2}} \right) \ , \tag{6}$$

where for any positive operator $\hat{A}$ we mean by $\hat{A}^{1/2}$ $\left(\text{or } \sqrt{\hat{A}}\right)$ the unique positive operator such that $\hat{A}^{1/2}\hat{A}^{1/2} = \hat{A}$. The quantity on the right hand side of Eq. (6) has appeared before in slightly different forms: as the distance function $d_{\mathrm{B}}^2(\hat{\rho}_0, \hat{\rho}_1) = 2 - 2\cos(B(\hat{\rho}_0, \hat{\rho}_1))$ of Bures [1, 21], the generalized transition probability for mixed states $\mathrm{prob}(\hat{\rho}_0 \to \hat{\rho}_1) = \cos^2(B(\hat{\rho}_0, \hat{\rho}_1))$ of Uhlmann [2], and (in the same form as Uhlmann's) Jozsa's criterion [3] for fidelity of signals in a quantum communication channel. Moreover, in a roundabout way through the mathematical-physics literature (cf., for instance, in logical order [12], [22], [23], [24], and [2]) one can put together a result quite similar in spirit to Eq. (6)—that is, a maximization like (5) but, instead of over all POVMs, restricted to orthogonal projection valued measures. What is novel here is the explicit statistical interpretation, the simplicity and generality of the derivation, and the fact that it pinpoints the measurement by which Eq. (6) is attained. The method of choice in deriving Eq. (6) is an application of the Schwarz inequality in such a way that its specific conditions for equality can be met by a suitable measurement. This is of use here because the problem of maximizing the Bhattacharyya-Wootters distance is equivalent to simply minimizing its cosine; that is, to prove Eq. (6), we need to show that

$$\min_{\{\hat{E}_b\}} \sum_b \sqrt{\mathrm{tr}\hat{\rho}_0\hat{E}_b}\sqrt{\mathrm{tr}\hat{\rho}_1\hat{E}_b} \ = \ \mathrm{tr}\sqrt{\hat{\rho}_1^{1/2}\hat{\rho}_0\hat{\rho}_1^{1/2}} \ . \tag{7}$$

First, however, it is instructive to consider a quick and dirty, and for this problem inappropriate, application of the Schwarz inequality; the difficulties encountered therein point naturally toward the correct proof. The Schwarz inequality for the operator inner product $\mathrm{tr}(\hat{A}^\dagger\hat{B})$ is given by $|\mathrm{tr}(\hat{A}^\dagger\hat{B})|^2 \leq \mathrm{tr}(\hat{A}^\dagger\hat{A})\mathrm{tr}(\hat{B}^\dagger\hat{B})$, where equality is achieved if and only

if $\hat{B} = \mu\hat{A}$ for some constant μ. Let $\{\hat{E}_b\}$ be an arbitrary POVM, $p_{0b} = \mathrm{tr}(\hat{\rho}_0\hat{E}_b)$, and $p_{1b} = \mathrm{tr}(\hat{\rho}_1\hat{E}_b)$. By the cyclic property of the trace and this inequality, we must have for any b,

$$\begin{aligned}\sqrt{p_{0b}}\sqrt{p_{1b}} &= \sqrt{\mathrm{tr}\Big(\big(\hat{E}_b^{1/2}\hat{\rho}_0^{1/2}\big)^\dagger\big(\hat{E}_b^{1/2}\hat{\rho}_0^{1/2}\big)\Big)}\sqrt{\mathrm{tr}\Big(\big(\hat{E}_b^{1/2}\hat{\rho}_1^{1/2}\big)^\dagger\big(\hat{E}_b^{1/2}\hat{\rho}_1^{1/2}\big)\Big)} \\ &\geq \left|\mathrm{tr}\Big(\big(\hat{E}_b^{1/2}\hat{\rho}_0^{1/2}\big)^\dagger\big(\hat{E}_b^{1/2}\hat{\rho}_1^{1/2}\big)\Big)\right| = \left|\mathrm{tr}\big(\hat{\rho}_0^{1/2}\hat{E}_b\hat{\rho}_1^{1/2}\big)\right| . \qquad (8)\end{aligned}$$

The condition for attaining equality here is that

$$\hat{E}_b^{1/2}\hat{\rho}_1^{1/2} = \mu_b\hat{E}_b^{1/2}\hat{\rho}_0^{1/2} . \qquad (9)$$

A subscript b has been placed on the constant μ as a reminder of its dependence on the particular $\hat{E}_b$ in this equation. From inequality (8), it follows by the linearity of the trace and the completeness property of POVMs that

$$\sum_b \sqrt{p_{0b}}\sqrt{p_{1b}} \geq \sum_b \left|\mathrm{tr}\big(\hat{\rho}_0^{1/2}\hat{E}_b\hat{\rho}_1^{1/2}\big)\right| \geq \left|\sum_b \mathrm{tr}\big(\hat{\rho}_0^{1/2}\hat{E}_b\hat{\rho}_1^{1/2}\big)\right| = \mathrm{tr}\big(\hat{\rho}_0^{1/2}\hat{\rho}_1^{1/2}\big) . \qquad (10)$$

The quantity $\mathrm{tr}\big(\hat{\rho}_0^{1/2}\hat{\rho}_1^{1/2}\big)$ is thus a lower bound to $\cos(B(p_0, p_1))$; for it to actually be the minimum, there must be a POVM such that, for all b, Eq. (9) is satisfied and $\mathrm{tr}\big(\hat{\rho}_0^{1/2}\hat{E}_b\hat{\rho}_1^{1/2}\big)$ is real and non-negative (from Eq. (10)). These conditions, though, cannot be fulfilled by any POVM $\{\hat{E}_b\}$, except in the case that $\hat{\rho}_0$ and $\hat{\rho}_1$ commute. This can be seen as follows. Suppose $[\hat{\rho}_0, \hat{\rho}_1] \neq 0$. Since $\hat{\rho}_0$ can be inverted, condition (9) can be written equivalently as

$$\hat{E}_b^{1/2}\big(\mu_b\hat{\mathbb{1}} - \hat{\rho}_1^{1/2}\hat{\rho}_0^{-1/2}\big) = 0 . \qquad (11)$$

The only way this can be satisfied is if we take the $\hat{E}_b$ to be proportional to the projectors formed from the *left*-eigenvectors of $\hat{\rho}_1^{1/2}\hat{\rho}_0^{-1/2}$ and let the μ_b be the corresponding eigenvalues. This is seen easily. The operator $\hat{\rho}_1^{1/2}\hat{\rho}_0^{-1/2}$ is a non-Hermitian operator on an n-dimensional Hilbert space, say, and thus has n linearly independent but non-orthogonal left-eigenvectors $\langle\psi_r|$ with eigenvalues σ_r and n linearly independent but non-orthogonal right-eigenvectors $|\phi_q\rangle$ with eigenvalues λ_q. Consider the operation of $\hat{E}_b^{1/2}\big(\mu_b\hat{\mathbb{1}} - \hat{\rho}_1^{1/2}\hat{\rho}_0^{-1/2}\big)$ on $|\phi_q\rangle$. Equation (11) implies $(\mu_b - \lambda_q)\hat{E}_b^{1/2}|\phi_q\rangle = 0$ for all q and b. Assume now, for simplicity, that all the λ_q are distinct. If $\hat{E}_b$ is not to be identically zero, then we must have that (modulo relabeling) $\hat{E}_b^{1/2}|\phi_q\rangle = 0$ for all $q \neq b$ and $\mu_b = \lambda_q$ for $q = b$. This means that $\hat{E}_b^{1/2}$ is proportional to the projector onto the one-dimensional subspace that is orthogonal to all the $|\phi_q\rangle$ for $q \neq b$. But since $0 = \langle\psi_r|\hat{\rho}_1^{1/2}\hat{\rho}_0^{-1/2}|\phi_q\rangle - \langle\psi_r|\hat{\rho}_1^{1/2}\hat{\rho}_0^{-1/2}|\phi_q\rangle = (\sigma_r - \lambda_q)\langle\psi_r|\phi_q\rangle$, we have that (again modulo relabeling) $|\psi_r\rangle$ is orthogonal to $|\phi_q\rangle$ for $q \neq r$ and $\sigma_r = \lambda_q$ for $q = r$, and therefore $\hat{E}_b^{1/2} \propto |\psi_b\rangle\langle\psi_b|$. The reason Eq. (11) cannot be satisfied by *any* POVM is just that the $|\psi_b\rangle$ are non-orthogonal. When the $|\psi_b\rangle$ are non-orthogonal, there are *no* positive constants α_b $(b = 1, \ldots, n)$ such that $\sum_b \alpha_b|\psi_b\rangle\langle\psi_b| = \hat{\mathbb{1}}$.

The lesson from this example is that the naïve Schwarz inequality is not enough to prove Eq. (6); one must be careful to "build in" a way to attain equality by at least one POVM.

Plainly the way to do this is to take advantage of the invariances of the trace operation. In particular, in the set of inequalities (8) we could have first written

$$\mathrm{tr}(\hat{\rho}_0\hat{E}_b) = \mathrm{tr}\Big(\hat{U}\hat{\rho}_0^{1/2}\hat{E}_b\hat{\rho}_0^{1/2}\hat{U}^\dagger\Big) \tag{12}$$

for any unitary operator $\hat{U}$. Then, in the same manner as there, it follows that

$$\sqrt{p_{0b}}\sqrt{p_{1b}} \geq \Big|\mathrm{tr}\Big(\hat{U}\hat{\rho}_0^{1/2}\hat{E}_b\hat{\rho}_1^{1/2}\Big)\Big| \ , \tag{13}$$

where the condition for equality is now $\hat{E}_b^{1/2}\hat{\rho}_1^{1/2} = \mu_b\hat{E}_b^{1/2}\hat{\rho}_0^{1/2}\hat{U}^\dagger$, which, because $\hat{\rho}_1$ is invertible, is equivalent to

$$\hat{E}_b^{1/2}\Big(\hat{\mathbb{1}} - \mu_b\hat{\rho}_0^{1/2}\hat{U}^\dagger\hat{\rho}_1^{-1/2}\Big) = 0 \ . \tag{14}$$

Finally, in the manner of Eq. (10), we get

$$\sum_b \sqrt{p_{0b}}\sqrt{p_{1b}} \geq \Big|\mathrm{tr}\Big(\hat{U}\hat{\rho}_0^{1/2}\hat{\rho}_1^{1/2}\Big)\Big| \ . \tag{15}$$

The condition for equality in this is to satisfy both Eq. (14) and the requirement that $\mathrm{tr}\Big(\hat{U}\hat{\rho}_0^{1/2}\hat{E}_b\hat{\rho}_1^{1/2}\Big)$ be real and non-negative for all b. Just as in the last example, though, there can be no POVM $\{\hat{E}_b\}$ that satisfies condition (14) *unless* the operator $\hat{\rho}_0^{1/2}\hat{U}^\dagger\hat{\rho}_1^{-1/2}$ is Hermitian (so that its eigenvectors form a complete orthonormal basis). An easy way to find a unitary $\hat{U}$ that makes a valid solution to Eq. (14) possible is to note a completely different point about inequality (15). The unitary operator $\hat{U}$ there is arbitrary; if there is to be a chance of attaining equality in (15), $\hat{U}$ had better be chosen so as to maximize $\Big|\mathrm{tr}\Big(\hat{U}\hat{\rho}_0^{1/2}\hat{\rho}_1^{1/2}\Big)\Big|$. It turns out that that particular $\hat{U}$ forces $\hat{\rho}_0^{1/2}\hat{U}^\dagger\hat{\rho}_1^{-1/2}$ to be Hermitian.

To demonstrate the last point, we rely on a result from the mathematical literature [3, 25, 26]: for any operator $\hat{A}$, $\max_{\hat{U}}\Big|\mathrm{tr}\Big(\hat{U}\hat{A}\Big)\Big| = \mathrm{tr}\sqrt{\hat{A}^\dagger\hat{A}}$, where the maximum is taken over all unitary operators $\hat{U}$; the particular $\hat{U}$ that gives rise to the maximum is defined by $\hat{U}\hat{A} = \sqrt{\hat{A}^\dagger\hat{A}}$. Thus the $\hat{U}$ that gives rise to the tightest inequality in Eq. (15) is

$$\hat{U}_{\mathrm{c}} \equiv \sqrt{\hat{\rho}_1^{1/2}\hat{\rho}_0\hat{\rho}_1^{1/2}}\,\hat{\rho}_1^{-1/2}\hat{\rho}_0^{-1/2} \ , \tag{16}$$

so that Eq. (15) now takes the form needed to prove Eq. (7):

$$\sum_b \sqrt{p_{0b}}\sqrt{p_{1b}} \geq \Big|\mathrm{tr}\Big(\hat{U}_{\mathrm{c}}\,\hat{\rho}_0^{1/2}\hat{\rho}_1^{1/2}\Big)\Big| = \mathrm{tr}\,\sqrt{\hat{\rho}_1^{1/2}\hat{\rho}_0\hat{\rho}_1^{1/2}} \ . \tag{17}$$

Inserting this choice for $\hat{U}$ into Eq. (14) gives the condition

$$\hat{E}_b^{1/2}\Big(\hat{\mathbb{1}} - \mu_b\hat{\rho}_1^{-1/2}\sqrt{\hat{\rho}_1^{1/2}\hat{\rho}_0\hat{\rho}_1^{1/2}}\,\hat{\rho}_1^{-1/2}\Big) = 0 \ . \tag{18}$$

The operator

$$\hat{M} \equiv \hat{\rho}_1^{-1/2}\sqrt{\hat{\rho}_1^{1/2}\hat{\rho}_0\hat{\rho}_1^{1/2}}\,\hat{\rho}_1^{-1/2} \tag{19}$$

in this equation is indeed Hermitian and also non-negative (as can be seen immediately from its symmetry). Thus there is a POVM $\{\hat{E}_b^{\rm B}\}$ that satisfies Eq. (14) for each b: the $\hat{E}_b^{\rm B}$ are just the projectors onto a basis that diagonalizes $\hat{M}$. Here the μ_b must be taken to be reciprocals of $\hat{M}$'s eigenvalues.

With the POVM $\{\hat{E}_b^{\rm B}\}$, the further condition that $\mathrm{tr}\big(\hat{U}_{\rm c}\,\hat{\rho}_0^{1/2}\hat{E}_b\hat{\rho}_1^{1/2}\big)$ be real and non-negative is automatically satisfied. Since the eigenvalues $1/\mu_b$ of $\hat{M}$ are all non-negative, one finds that

$$\mathrm{tr}\big(\hat{U}_{\rm c}\,\hat{\rho}_0^{1/2}\hat{E}_b^{\rm B}\hat{\rho}_1^{1/2}\big) = \mathrm{tr}\big(\hat{\rho}_1\hat{M}\hat{E}_b^{\rm B}\big) = \frac{1}{\mu_b}\mathrm{tr}\big(\hat{\rho}_1\hat{E}_b^{\rm B}\big) \geq 0\ . \tag{20}$$

This concludes the proof of Eq. (6): the Bhattacharyya-Wootters distance maximized over *all* quantum measurements is a simple function of Uhlmann's transition probability.

In the remainder of this section we report a few interesting points about the measurement specified by $\hat{M}$ and the quantum distinguishability measure $B(\hat{\rho}_0,\hat{\rho}_1)$. Equation (2) defining the Bhattacharyya-Wootters distance is clearly invariant under interchanges of the labels 0 and 1. Therefore it must follow that $B(\hat{\rho}_0,\hat{\rho}_1) = B(\hat{\rho}_1,\hat{\rho}_0)$. A neat way to see this directly is to note that the operators $\hat{\rho}_1^{1/2}\hat{\rho}_0\hat{\rho}_1^{1/2}$ and $\hat{\rho}_0^{1/2}\hat{\rho}_1\hat{\rho}_0^{1/2}$ have the same eigenvalue spectrum. For if $|b\rangle$ and λ_b are an eigenvector and eigenvalue of $\hat{\rho}_1^{1/2}\hat{\rho}_0\hat{\rho}_1^{1/2}$, it follows that

$$\lambda_b\big(\hat{\rho}_0^{1/2}\hat{\rho}_1^{1/2}|b\rangle\big) = \hat{\rho}_0^{1/2}\hat{\rho}_1^{1/2}\big(\lambda_b|b\rangle\big) = \hat{\rho}_0^{1/2}\hat{\rho}_1^{1/2}\big(\hat{\rho}_1^{1/2}\hat{\rho}_0\hat{\rho}_1^{1/2}|b\rangle\big) = \big(\hat{\rho}_0^{1/2}\hat{\rho}_1\hat{\rho}_0^{1/2}\big)\big(\hat{\rho}_0^{1/2}\hat{\rho}_1^{1/2}|b\rangle\big)\ .$$

Hence, $\mathrm{tr}\,\sqrt{\hat{\rho}_1^{1/2}\hat{\rho}_0\hat{\rho}_1^{1/2}} = \mathrm{tr}\,\sqrt{\hat{\rho}_0^{1/2}\hat{\rho}_1\hat{\rho}_0^{1/2}}$ and so $B(\hat{\rho}_0,\hat{\rho}_1) = B(\hat{\rho}_1,\hat{\rho}_0)$. By the same token, the derivation of Eq. (6) itself must remain valid if all the 0's and 1's in it are interchanged throughout. This, however, would give rise to a measurement specified by a basis diagonalizing $\hat{N} \equiv \hat{\rho}_0^{-1/2}\sqrt{\hat{\rho}_0^{1/2}\hat{\rho}_1\hat{\rho}_0^{1/2}}\hat{\rho}_0^{-1/2}$. It turns out that $\hat{M}$ and $\hat{N}$ can define the same measurement because not only do they commute, they are inverses of each other. This can be seen as follows. Let $\hat{A}$ be any operator and $\hat{V}$ be a unitary operator such that $\hat{A} = \hat{V}\sqrt{\hat{A}^\dagger\hat{A}}$ and hence $\sqrt{\hat{A}^\dagger\hat{A}} = \hat{V}^\dagger\hat{A}$. Then $\hat{A}^\dagger = \sqrt{\hat{A}^\dagger\hat{A}}\,\hat{V}^\dagger$ and $\big(\hat{V}\sqrt{\hat{A}^\dagger\hat{A}}\,\hat{V}^\dagger\big)^2 = \big(\hat{V}\sqrt{\hat{A}^\dagger\hat{A}}\big)\big(\sqrt{\hat{A}^\dagger\hat{A}}\,\hat{V}^\dagger\big) = \hat{A}\hat{A}^\dagger$, and therefore $\sqrt{\hat{A}\hat{A}^\dagger} = \hat{V}\sqrt{\hat{A}^\dagger\hat{A}}\,\hat{V}^\dagger = \hat{V}\hat{A}^\dagger$. In particular, if $\sqrt{\hat{\rho}_1^{1/2}\hat{\rho}_0\hat{\rho}_1^{1/2}} = \hat{U}_{\rm c}\,\hat{\rho}_0^{1/2}\hat{\rho}_1^{1/2}$, then $\sqrt{\hat{\rho}_0^{1/2}\hat{\rho}_1\hat{\rho}_0^{1/2}} = \hat{U}_{\rm c}^\dagger\,\hat{\rho}_1^{1/2}\hat{\rho}_0^{1/2}$ and hence

$$\hat{M}\hat{N} = \hat{\rho}_1^{-1/2}\big(\hat{U}_{\rm c}\,\hat{\rho}_0^{1/2}\hat{\rho}_1^{1/2}\big)\hat{\rho}_1^{-1/2}\hat{\rho}_0^{-1/2}\big(\hat{U}_{\rm c}^\dagger\,\hat{\rho}_1^{1/2}\hat{\rho}_0^{1/2}\big)\hat{\rho}_0^{-1/2} = \hat{\rho}_1^{-1/2}\hat{U}_{\rm c}\,\hat{U}_{\rm c}^\dagger\,\hat{\rho}_1^{1/2} = \hat{\mathbb{1}}\ . \tag{21}$$

Finally, we note an interesting expression for $\hat{M}$'s eigenvalues that arises from the last result. Let the eigenvalues and eigenvectors of $\hat{M}$ be denoted by m_b and $|b\rangle$; in this notation $\hat{E}_b^{\rm B} = |b\rangle\langle b|$. Then we can write two expressions for m_b:

$$m_b\langle b|\hat{\rho}_1|b\rangle = \langle b|\hat{\rho}_1\hat{M}|b\rangle = \langle b|\hat{\rho}_1^{1/2}\hat{U}_{\rm c}\,\hat{\rho}_0^{1/2}|b\rangle\ , \tag{22}$$

$$\frac{1}{m_b}\langle b|\hat{\rho}_0|b\rangle = \langle b|\hat{\rho}_0\hat{N}|b\rangle = \langle b|\hat{\rho}_0^{1/2}\hat{U}_{\rm c}^\dagger\,\hat{\rho}_1^{1/2}|b\rangle = \big(\langle b|\hat{\rho}_1^{1/2}\hat{U}_{\rm c}\,\hat{\rho}_0^{1/2}|b\rangle\big)^*\ . \tag{23}$$

Because the left hand sides of these equations are real numbers, so are the right hand sides; in particular, combining Eqs. (22) and (23), we get

$$m_b = \left(\frac{\langle b|\hat{\rho}_0|b\rangle}{\langle b|\hat{\rho}_1|b\rangle}\right)^{1/2} = \left(\frac{\mathrm{tr}\hat{\rho}_0\hat{E}_b^{\rm B}}{\mathrm{tr}\hat{\rho}_1\hat{E}_b^{\rm B}}\right)^{1/2}\ . \tag{24}$$

Thus the optimal measurement operator $\hat{M}$ for the Bhattacharyya-Wootters distance might be considered a sort of operator analog to the classical likelihood ratio. This fact gives rise to an interesting expression for the Kullback-Leibler relative information (1) between $\hat{\rho}_0$ and $\hat{\rho}_1$ with respect to this measurement:

$$\begin{aligned} K_{\rm B}(\hat{\rho}_0/\hat{\rho}_1) &\equiv \sum_b \Big({\rm tr}\hat{\rho}_0\hat{E}_b^{\rm B}\Big)\ln\left(\frac{{\rm tr}\hat{\rho}_0\hat{E}_b^{\rm B}}{{\rm tr}\hat{\rho}_1\hat{E}_b^{\rm B}}\right) = 2\,{\rm tr}\left(\hat{\rho}_0\sum_b(\ln\, m_b)\hat{E}_b^{\rm B}\right) \\ &= 2\,{\rm tr}\left(\hat{\rho}_0\ln\left(\hat{\rho}_1^{-1/2}\sqrt{\hat{\rho}_1^{1/2}\hat{\rho}_0\hat{\rho}_1^{1/2}}\hat{\rho}_1^{-1/2}\right)\right)\ . \end{aligned} \tag{25}$$

This, of course, will generally *not* be the maximum of the Kullback-Leibler information over all measurements, but it does provide a lower bound for the maximum value. Moreover, a quantity quite similar to this arises naturally in the context of still another measure of quantum distinguishability studied by Braunstein and Caves [19].

3 Accessible Information

A binary quantum communication channel is defined by its signal states $\{\hat{\rho}_0, \hat{\rho}_1\}$ and their prior probabilities $\{1-t, t\}$ $(0 \le t \le 1)$. The Shannon mutual information [27] for the channel with respect to a measurement $\{\hat{E}_b\}$ is

$$I(t) \equiv H(p) - (1-t)H(p_0) - tH(p_1) = (1-t)K(p_0/p) + tK(p_1/p)\ , \tag{26}$$

where $H(p) = -\sum_b p_b \ln p_b$ is the Shannon information of the probability distribution p_b, $p_{0b} = {\rm tr}(\hat{\rho}_0\hat{E}_b)$, $p_{1b} = {\rm tr}(\hat{\rho}_1\hat{E}_b)$, and $p_b = {\rm tr}(\hat{\rho}\hat{E}_b)$ for $\hat{\rho} = (1-t)\hat{\rho}_0 + t\hat{\rho}_1 = \hat{\rho}_0 + t\hat{\Delta} = \hat{\rho}_1 - (1-t)\hat{\Delta}$ with $\hat{\Delta} = \hat{\rho}_1 - \hat{\rho}_0$. The *accessible information* $I_{\rm acc}(t)$ is the mutual information $I(t)$ maximized over all measurements $\{\hat{E}_b\}$.

The problems associated with actually finding $I_{\rm acc}(t)$ and the measurement that gives rise to it are every bit as difficult as those in maximizing the Kullback-Leibler information, perhaps more so—for here it is not only the logarithm that confounds things, but also the fact that $\hat{\rho}_0$ and $\hat{\rho}_1$ are "coupled" through the mean density operator $\hat{\rho}$. There does, at least, exist a general upper bound to $I_{\rm acc}(t)$, due to Holevo [4], but that is of little use in pinpointing the measurement that gives rise to $I_{\rm acc}(t)$. In what follows, we simplify the derivation of the Holevo bound via a variation of the methods used in the last section. This simplification has the advantage of specifying a measurement whose use lower bounds $I_{\rm acc}(t)$.

The Holevo upper bound to $I_{\rm acc}(t)$ is

$$I_{\rm acc}(t) \le S(\hat{\rho}) - (1-t)S(\hat{\rho}_0) - tS(\hat{\rho}_1) \equiv S(t)\ , \tag{27}$$

where $S(\hat{\rho}) = -{\rm tr}(\hat{\rho}\ln\hat{\rho}) = -\sum_j \lambda_j \ln\lambda_j$ is the von Neumann entropy of the density operator $\hat{\rho}$, whose eigenvalues are λ_j. The key to deriving it is in realizing the importance of properties of $I(t)$ and $S(t)$ as functions of t [4]. Note that $I(0) = I(1) = S(0) = S(1) = 0$. Moreover, both $I(t)$ and $S(t)$ are downwardly convex, as can be seen by working out their second derivatives. For $I(t)$ a straightforward calculation gives

$$I''(t) = -\sum_b \frac{\Big({\rm tr}(\hat{\Delta}\hat{E}_b)\Big)^2}{{\rm tr}(\hat{\rho}\hat{E}_b)}\ . \tag{28}$$

For $S(t)$ it is easiest to proceed by representing $S(\hat{\rho})$ as a contour integral [28]

$$S(\hat{\rho}) = -\frac{1}{2\pi i}\oint_C (z\ln z)\ \mathrm{tr}\Big((z\hat{\mathbb{1}} - \hat{\rho})^{-1}\Big)\, dz\ , \tag{29}$$

where the contour C encloses all the *nonzero* eigenvalues of $\hat{\rho}$; by differentiating within the integral and using the operator identity $(\hat{A}^{-1})' = -\hat{A}^{-1}\hat{A}'\hat{A}^{-1}$, one finds that

$$S''(t) = -\sum_{\{j,k|\lambda_j+\lambda_k\neq 0\}} \Phi(\lambda_j,\lambda_k)|\Delta_{jk}|^2\ , \tag{30}$$

where $\Phi(x,y) = (\ln x - \ln y)/(x-y)$ if $x \neq y$, $\Phi(x,x) = 1/x$, $\Delta_{jk} = \langle j|\hat{\Delta}|k\rangle$, and $|j\rangle$ is the eigenvector of $\hat{\rho}$ with eigenvalue λ_j. Expressions (28) and (30) are clearly non-positive.

The statement that $S(t)$ is an upper bound to $I(t)$ for *any* t is equivalent to the property that, when plotted versus t, the curve for $S(t)$ has a more negative curvature than the curve for $I(t)$ (regardless of which POVM $\{\hat{E}_b\}$ is used in its definition), i.e., $S''(t) \leq I''(t) \leq 0$ for any POVM $\{\hat{E}_b\}$. The meat of the derivation is in showing this inequality. Holevo does this by demonstrating the existence of a function $L''(t)$, independent of $\{\hat{E}_b\}$, such that $S''(t) \leq L''(t)$ and $L''(t) \leq I''(t)$. From this it follows, upon enforcing the boundary condition $L(0) = L(1) = 0$, that $I_{acc}(t) \leq L(t) \leq S(t)$.

It is at this point that a fairly drastic simplification can be made to the original proof. An easy way to get at such a function $L''(t)$ is simply to minimize $I''(t)$ over all POVMs $\{\hat{E}_b\}$, and thereafter to show that $S''(t) \leq L''(t)$. This, again, is distinctly more tractable than extremizing the mutual information $I(t)$ itself because no logarithms appear in $I''(t)$; there is hope for solution by means of the Schwarz inequality. This approach, it turns out, generates exactly the same function $L''(t)$ as used by Holevo in the original proof, though the two derivations appear to have little to do with each other. The difference of importance here is that this approach pinpoints the measurement that actually minimizes $I''(t)$. This measurement, though it generally does not maximize $I(t)$ itself, necessarily does provide a *lower* bound $M(t)$ to the accessible information $I_{\mathrm{acc}}(t)$ [29].

The problem of minimizing Eq. (28) is formally identical to the problem considered by Braunstein and Caves [17]: the expression for $-I''(t)$ is just the Fisher information of Eq. (3). The steps are as follows. The idea is to think of the numerator within the sum (28) as analogous to the left hand side of the Schwarz inequality. One would like to use the Schwarz inequality in such a way that the $\mathrm{tr}(\hat{\rho}\hat{E}_b)$ term in the denominator is cancelled and only an expression linear in $\hat{E}_b$ is left; for then, upon summing over the index b, the completeness property for POVMs will leave the final expression independent of the given measurement.

This can be done (at the very least) by introducing a "lowering" super-operator $\mathcal{G}_{\hat{C}}$ with the property that for any positive operators $\hat{A}$, $\hat{B}$, $\hat{C}$,

$$\mathrm{tr}(\hat{A}\hat{B}) \leq \Big|\mathrm{tr}\Big(\hat{C}\hat{B}\,\mathcal{G}_{\hat{C}}(\hat{A})\Big)\Big|\ . \tag{31}$$

For then one can derive by simple applications of the Schwarz inequality (just as in Eq. (8))

$$\Big(\mathrm{tr}(\hat{\Delta}\hat{E}_b)\Big)^2 \leq \Big|\mathrm{tr}\left(\hat{\rho}\hat{E}_b\mathcal{G}_{\hat{\rho}}(\hat{\Delta})\right)\Big|^2 \leq \Big(\mathrm{tr}\hat{\rho}\hat{E}_b\Big)\mathrm{tr}\Big(\hat{E}_b\mathcal{G}_{\hat{\rho}}(\hat{\Delta})\hat{\rho}\,\mathcal{G}_{\hat{\rho}}(\hat{\Delta})^\dagger\Big) \tag{32}$$

and

$$\Big(\mathrm{tr}(\hat{\Delta}\hat{E}_b)\Big)^2 \leq \Big|\mathrm{tr}\Big(\hat{\rho}^{1/2}\hat{E}_b\mathcal{G}_{\hat{\rho}^{1/2}}(\hat{\Delta})\Big)\Big|^2 \leq \Big(\mathrm{tr}\hat{\rho}\hat{E}_b\Big)\mathrm{tr}\Big(\hat{E}_b\mathcal{G}_{\hat{\rho}^{1/2}}(\hat{\Delta})\,\mathcal{G}_{\hat{\rho}^{1/2}}(\hat{\Delta})^\dagger\Big)\ , \tag{33}$$

where the conditions for equality in the rightmost inequalities of these are, respectively,

$$\hat{E}_b^{1/2}\mathcal{G}_{\hat{\rho}}(\hat{\Delta})\hat{\rho}^{1/2} = \mu_b\hat{E}_b^{1/2}\hat{\rho}^{1/2} \qquad \text{and} \qquad \hat{E}_b^{1/2}\mathcal{G}_{\hat{\rho}^{1/2}}(\hat{\Delta}) = \mu_b\hat{E}_b^{1/2}\hat{\rho}^{1/2}\ . \tag{34}$$

Using inequalities (32) and (33) in Eq. (28) for $I''(t)$ immediately gives the lower bounds

$$I''(t) \geq -\mathrm{tr}\Big(\mathcal{G}_{\hat{\rho}}(\hat{\Delta})\hat{\rho}\,\mathcal{G}_{\hat{\rho}}(\hat{\Delta})^\dagger\Big) \qquad \text{and} \qquad I''(t) \geq -\mathrm{tr}\Big(\mathcal{G}_{\hat{\rho}^{1/2}}(\hat{\Delta})\,\mathcal{G}_{\hat{\rho}^{1/2}}(\hat{\Delta})^\dagger\Big)\ . \tag{35}$$

The problem now, much like in the last section, is to choose a super-operator $\mathcal{G}_{\hat{\rho}}$ in such a way that equality can be attained in Eq. (35). The "lowering" super-operator $\mathcal{L}_{\hat{\rho}}$ that does the trick [17] is defined by its action on an operator $\hat{A}$ by

$$\frac{1}{2}\Big(\hat{\rho}\mathcal{L}_{\hat{\rho}}(\hat{A}) + \mathcal{L}_{\hat{\rho}}(\hat{A})\hat{\rho}\Big) = \hat{A}. \tag{36}$$

In a basis $|j\rangle$ that diagonalizes $\hat{\rho}$, $\mathcal{L}_{\hat{\rho}}(\hat{\Delta})$ becomes

$$\mathcal{L}_{\hat{\rho}}(\hat{\Delta}) \equiv \sum_{\{j,k|\lambda_j+\lambda_k\neq 0\}} \frac{2}{\lambda_j+\lambda_k}\Delta_{jk}|j\rangle\langle k|\ , \tag{37}$$

which depends on the fact that $\Delta_{jk} = 0$ if $\lambda_j + \lambda_k = 0$. (For further discussion of why Eq. (37) is the appropriate extension of $\mathcal{L}_{\hat{\rho}}(\hat{A})$ to the zero-eigenvalue subspaces of $\hat{\rho}$, see [17]; note that $\mathcal{L}_{\hat{\rho}}$ is denoted there by $\mathcal{R}_{\hat{\rho}}^{-1}$.) This super-operator is easily seen, using Eq. (36), to satisfy the identity that for Hermitian $\hat{A}$ and $\hat{B}$, $\mathrm{tr}(\hat{A}\hat{B}) = \mathrm{Re}[\mathrm{tr}(\hat{\rho}\hat{A}\mathcal{L}_{\hat{\rho}}(\hat{B}))]$ and, hence, also to satisfy Eq. (31). The desired optimization is via the left member of Eq. (35):

$$I''(t) \geq -\mathrm{tr}(\mathcal{L}_{\hat{\rho}}(\hat{\Delta})\hat{\rho}\mathcal{L}_{\hat{\rho}}(\hat{\Delta})) = -\mathrm{tr}(\hat{\Delta}\mathcal{L}_{\hat{\rho}}(\hat{\Delta})) = -\sum_{\{j,k|\lambda_j+\lambda_k\neq 0\}} \frac{2}{\lambda_j+\lambda_k}\,|\Delta_{jk}|^2\ . \tag{38}$$

The conditions for equality in Eq. (38) are $\mathrm{Im}[\mathrm{tr}(\hat{\rho}\hat{E}_b\mathcal{L}_{\hat{\rho}}(\hat{\Delta}))] = 0$ for all b and (from Eq. (34))

$$\hat{E}_b^{1/2}\Big(\mu_b\hat{\mathbb{1}} - \mathcal{L}_{\hat{\rho}}(\hat{\Delta})\Big) = 0 \quad \text{for all } b\ . \tag{39}$$

Both conditions can always be met by choosing the operators $\hat{E}_b^{\mathrm{F}}$ to be projectors onto the basis that diagonalizes the Hermitian operator $\mathcal{L}_{\hat{\rho}}(\hat{\Delta})$ and choosing the constants μ_b to be the eigenvalues of $\mathcal{L}_{\hat{\rho}}(\hat{\Delta})$.

The function $L''(t)$ can now be defined as $L''(t) = -\mathrm{tr}(\hat{\Delta}\mathcal{L}_{\hat{\rho}}(\hat{\Delta}))$. This, as stated above, is exactly the function $L''(t)$ used by Holevo, but obtained there by other means. The remainder of the derivation of Eq. (27), to show that $S''(t) \leq L''(t)$, consists of demonstrating the arithmetic inequality $\Phi(x,y) \geq 2/(x+y)$ (see [4]).

Finally we focus on deriving an explicit expression for the lower bound $M(t)$. In the manner of Eq. (25) the mutual information can be written as

$$I(t) = \mathrm{tr}\bigg((1-t)\hat{\rho}_0\sum_b(\ln\alpha_b)\hat{E}_b + \ t\,\hat{\rho}_1\sum_b(\ln\beta_b)\hat{E}_b\bigg)\,, \tag{40}$$

where $\alpha_b = \mathrm{tr}(\hat{\rho}_0\hat{E}_b)/\mathrm{tr}(\hat{\rho}\hat{E}_b)$ and $\beta_b = \mathrm{tr}(\hat{\rho}_1\hat{E}_b)/\mathrm{tr}(\hat{\rho}\hat{E}_b)$. The lower bound $M(t)$ is defined by inserting the projectors $\hat{E}_b^{\mathrm{F}}$ onto a basis that diagonalizes $\mathcal{L}_{\hat{\rho}}(\hat{\Delta})$ into this formula. Now

a curious fact can be used: even though $\hat{\rho}_0$ and $\hat{\rho}_1$ need not commute, $\mathcal{L}_{\hat{\rho}}(\hat{\Delta})$, $\mathcal{L}_{\hat{\rho}}(\hat{\rho}_0)$, and $\mathcal{L}_{\hat{\rho}}(\hat{\rho}_1)$ do all commute. This follows from the linearity of the $\mathcal{L}_{\hat{\rho}}$ super-operator: $\mathcal{L}_{\hat{\rho}}(\hat{\rho}_0) = \mathcal{L}_{\hat{\rho}}(\hat{\rho} - t\hat{\Delta}) = \hat{\mathbb{1}} - t\mathcal{L}_{\hat{\rho}}(\hat{\Delta})$ and $\mathcal{L}_{\hat{\rho}}(\hat{\rho}_1) = \mathcal{L}_{\hat{\rho}}(\hat{\rho} + (1-t)\hat{\Delta}) = \hat{\mathbb{1}} + (1-t)\mathcal{L}_{\hat{\rho}}(\hat{\Delta})$. Thus the same projectors $\hat{E}_b^{\mathrm{F}}$ that diagonalize $\mathcal{L}_{\hat{\rho}}(\hat{\Delta})$ also diagonalize $\mathcal{L}_{\hat{\rho}}(\hat{\rho}_0)$ and $\mathcal{L}_{\hat{\rho}}(\hat{\rho}_1)$. With this, it immediately follows from Eq. (36) that α_b and β_b are the respective eigenvalues of $\mathcal{L}_{\hat{\rho}}(\hat{\rho}_0)$ and $\mathcal{L}_{\hat{\rho}}(\hat{\rho}_1)$ corresponding to the projector $\hat{E}_b^{\mathrm{F}}$. Hence $M(t)$ takes the form

$$M(t) = \mathrm{tr}\Big((1-t)\,\hat{\rho}_0 \ln\big(\mathcal{L}_{\hat{\rho}}(\hat{\rho}_0)\big) + t\,\hat{\rho}_1 \ln\big(\mathcal{L}_{\hat{\rho}}(\hat{\rho}_1)\big)\Big)\,. \tag{41}$$

Similarly, one can obtain another lower bound (distinct from Eq. (25)) to the maximum Kullback-Leibler information by using the measurment basis that diagonalizes $\mathcal{L}_{\hat{\rho}_1}(\hat{\rho}_0)$.

References

[∗] This work was supported in part by the Office of Naval Research (Grant. No. N00014-93-1-0116).

[1] D. J. C. Bures, Trans. Am. Math. Soc. **135**, 199 (1969).

[2] A. Uhlmann, Rep. Math. Phys. **9**, 273 (1976).

[3] R. Jozsa, "Fidelity for Mixed Quantum States," to appear in Journal of Modern Optics (1994).

[4] A. S. Holevo (Kholevo), Probl. Peredachi Inf. **9**(3), 3 (1973) [Prob. Inf. Transm. **9**, 177 (1973)].

[5] H. P. Yuen and M. Ozawa, Phys. Rev. Lett. **70**, 363 (1993).

[6] M. J. W. Hall and M. J. O'Rourke, Quantum Opt. **5**, 161 (1993).

[7] L. B. Levitin, in *Proceedings of the Fourth All-Union Conference on Information and Coding Theory*, Sec. II (Tahkent, 1969).

[8] S. Kullback and R. A. Leibler, Ann. Math. Stat. **22**, 79 (1951).

[9] T. Kailath, IEEE Trans. Commun. Technol. **COM-15**(1), 52 (1967).

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).

[11] A. Bhattacharyya, Bull. Calcutta Math. Soc. **35**, 99 (1943).

[12] W. K. Wootters, Phys. Rev. D **23**, 357 (1981).

[13] R. A. Fisher, Proc. R. Soc. Edinburgh **42**, 321 (1922).

[14] C. R. Rao, Sankhyā **9**, 246 (1949).

[15] S. Kullback, *Information Theory and Statistics* (Wiley, New York, 1959).

[16] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ, 1946).

[17] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).

[18] S. L. Braunstein, C. M. Caves, and G. J. Milburn, "Generalized Uncertainty Relations: Theory, Examples, and Lorentz Invariance," University of New Mexico Center for Advanced Studies preprint, submitted to Phys. Rev. A (1994).

[19] S. L. Braunstein and C. M. Caves, to appear in *Proceedings of the Conference on Quantum Communication and Measurement*, edited by R. Hudson, V. P. Belavkin, and O. Hirota (Plenum Press, New York, 1995), to be published.

[20] K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory* (Springer, Berlin, 1983).

[21] M. Hübner, Phys. Lett. A **163**, 239 (1992).

[22] N. Hadjisavvas, Commun. Math. Phys. **83**, 43 (1982).

[23] H. Araki and G. Raggio, Lett. Math. Phys. **6**, 237 (1982).

[24] N. Hadjisavvas, Lin. Alg. App. **84**, 281 (1986).

[25] K. Fan, Proc. Nat. Acad. Sci. **37**, 760 (1951).

[26] R. Schatten, *Norm Ideals of Completely Continuous Operators* (Springer-Verlag, Berlin, 1960).

[27] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 623 (1948).

[28] H. Poincaré, Trans. Cambridge Philos. Soc. **18**, 220 (1899).

[29] C. A. Fuchs and C. M. Caves, Phys. Rev. Lett. **73**, 3047 (1994).